NONNEGATIVE COMPRESSED SENSING WITH MINIMAL PERTURBED EXPANDERS

M.Amin Khajehnejad Alexandros G. Dimakis Babak Hassibi

Email: {amin,adim,hassibi}@caltech.edu

ABSTRACT

This paper studies compressed sensing for the recovery of *non-negative* sparse vectors from a smaller number of measurements than the ambient dimension of the unknown vector. We construct sparse measurement matrices for the recovery of non-negative vectors, using perturbations of adjacency matrices of expander graphs with much smaller expansion coefficients than previously suggested schemes. These constructions are crucial in applications, such as DNA microarrays and sensor networks, where dense measurements are not practically feasible. We present a necessary and sufficient condition for ℓ_1 optimization to successfully recover the unknown vector and obtain closed form expressions for the recovery threshold. We finally present a novel recovery algorithm that exploits expansion and is faster than ℓ_1 optimization.

Index Terms— compressed sensing, expander graph, non-negative vector, l_1 optimization, perfect matching.

1. INTRODUCTION

We investigate the problem of signal recovery in compressed sensing, i.e., the problem of reconstructing a signal $\mathbf{x}$ that is assumed to be k sparse using m measurements, $\mathbf{y} = \mathbf{A}\mathbf{x}$, where m is smaller than the ambient dimension of the signal n, but larger than k. $\mathbf{A}$ here is the $m \times n$ so-called measurement matrix. In this paper, we focus on the case where the nonzero entries of x are positive, a special case that is of great practical interest.

In compressed sensing, $\mathbf{A}$ is often a dense matrix drawn from some ensemble of random matrices (see, e.g., [3]). In this paper, however, we will focus on sparse measurement matrices. This is important for numerous reasons. In several applications, like DNA micro arrays, the cost of each measurement increases with the number of coordinates of $\mathbf{x}$ involved [15]. Also, sparse measurement matrices often make possible the design of faster decoding algorithms (e.g., [10, 6, 7, 18]) apart from the general linear programming decoder [3]. In addition, unlike random measurement matrices (such as Gaussian or Bernoulli), which only guarantee the recovery of sparse vectors with high probability, expander graphs give deterministic guarantees (see, e.g., [10], which gives a deterministic guarantee for the fast algorithm proposed, and [5] for concentration lemmas on expander graphs).

Unlike Gaussian matrices, where reasonably sharp bounds on the thresholds which guarantee linear programming to recover sparse signals have been obtained [2], such sharp bounds do not exist for expander-graph-based measurements. This is the main focus of the current paper, for the special case where the k-sparse vector is non-negative.

It turns out that, due to the additional non-negativity constraint, one requires significantly fewer measurements to recover k-sparse non-negative signals. The non-negative case has also been studied in [4] for Gaussian matrices and also in the work of Bruckstein et al. [9], which further proposes a "matching pursuit" type of recovery algorithm.

The success of a measurement matrices is often certified by a so-called Restricted Isometry Property (RIP) which guarantees the success of ℓ_1 minimization. Recently, Indyk et al [5] showed that the adjacency matrices of suitable unbalanced expander graphs satisfy an RIP property for $\ell_{p\sim 1}$ norm. However, it turns out that RIP conditions are only sufficient. A complete characterization of good measurement matrices was recently given in terms of their null space. As stated in previous work (e.g. [16, 19, 21, 23]). if for any vector $\mathbf{w}$ in the null space of $\mathbf{A}$, the sum of the absolute values of any k elements of $\mathbf{w}$ is less that the sum of the absolute values of the rest of the elements, then the solution to $\min ||\mathbf{x}||_0$ subject to $\mathbf{A}\mathbf{x} = \mathbf{y}$ can always be obtained by solving $\min ||\mathbf{x}||_1$ subject to $\mathbf{A}\mathbf{x} = \mathbf{y}$, provided $\mathbf{x}$ is k-sparse.[1] This condition is stated in the work of Donoho [1] as the k-neighborly polytope property of $\mathbf{A}$, and in the work of Candes et al. as the uncertainty principle [3]. Donoho et al. also have been able to show the validity of this condition with high probability for random i.i.d Gaussian matrices and are therefore able to compute fairly tight thresholds on when linear-programming-based compressed sensing works [2]. The first analysis of the null space for expander graphs has been done by Indyk [8], where it was shown that every $(2k,\epsilon)$ expander graph with $\epsilon \leq \frac{1}{6}$ will have a well supported null space. See also [17] for explicit constructions using expander graphs.

Using Theorem 1 of [13], which is a generalization of the null space property theorem for the recovery of approximately sparse signals, Indyk's result gives an upper bound on the error when linear programming is used to recover approximately k-sparse vectors using expander graph measurements.

[1] Here $\|\cdot\|_0$ represents the number of non-zero entries in its argument vector and $\|\cdot\|_1$ is the standard ℓ_1-norm.

2. PROBLEM FORMULATION

The goal in compressed sensing is to recover a sparse vector from a set of under-determined linear equations. In many real world applications the original data vector is nonnegative, which is the case that we will focus on in this paper. The original problem of compressed sensing for the nonnegative input vectors is as following:

$$\min_{\mathbf{Ax}=\mathbf{y},\mathbf{x}\geq 0} ||\mathbf{x}||_0 \tag{1}$$

where $\mathbf{A}^{m\times n}$ is the measurement matrix, $y^{m\times 1}$ is called the observation vector, $x^{n\times 1}$ is the unknown vector which is known to be k-sparse, i.e., to have only k nonzero entries, and where $\|\cdot\|_0$ is l_0 norm, i.e., the number of nonzero entries of a given vector. The typical situation is that $n > m > k$. Although (1) is an NP-hard problem, Donoho and Tanner have shown in [4] that for a class of matrices $\mathbf{A}$ maintaining a so-called outwardly k-neighborly property and $\mathbf{x}$ being at most k-sparse, the solution to (1) is unique and can be recovered via the following linear programming problem:

$$\min_{\mathbf{Ax}=\mathbf{y},\mathbf{x}\geq 0} ||\mathbf{x}||_1 \tag{2}$$

They also show that i.i.d Gaussian random $m \times n$ matrices with $m = n/2$ are outwardly $m/8$-neighborly with high probability, and thus allow the recovery of $n/16$ sparse vectors $\mathbf{x}$ via linear programming. They further define a *weak* neighborly notion, based upon which they show that the same Gaussian random matrices will allow the recovery of *almost* all $0.558m$ sparse vectors $\mathbf{x}$ via ℓ_1-optimization for sufficiently large n.

In this paper, we primarily seek the answer to a similar question when the measurement matrix $\mathbf{A}$ is the adjacency matrix of an unbalanced bipartite graph with constant left degree d. The aim is to analyze the outwardly neighborly conditions for this class of matrices and come up with sparse structures that allow the recovery of vectors with sparsity proportional to the number of equations.

3. NULL SPACE CHARACTERIZATION AND COMPLETE RANK

We begin by stating an equivalent version of the outwardly neighborly condition which is in fact similar to the null space property that was mentioned in the introduction, but for the non-negative case. Later we show that this has a much more mundane interpretation for the special case of regular bipartite graphs.

Theorem 3.1. *let* $\mathbf{A}$ *be a nonnegative* $m \times n$ *matrix and* $k < n/2$ *be a positive integer. These two statements are equivalent:*

- *For every nonnegative vector* $\mathbf{x}_0$ *with at most* k *nonzeros,* $\mathbf{x}_0$ *is a solution to (2) with* $\mathbf{y} = \mathbf{Ax}_0$.
- *For every vector* $\mathbf{w}$ *in the null space of* $\mathbf{A}$ *, and every index set* $S \subset \{1, 2, ..., n\}$ *with* $|S| = k$ *such that* $\mathbf{w}_{S^c}$ *is a* non-negative *vector,* $||\mathbf{w}_S||_1 \leq ||\mathbf{w}_{S^c}||_1$.[2]

Theorem 3.1 is in fact the counter part of Theorem 1 of [16] for nonnegative vectors and the proof is very similar. The theorem gives a necessary and sufficient condition on the matrix A, such that all k-sparse x_0 can be recovered using (2). The condition is essentially that for every vector in the null space of $\mathbf{A}$, the sum of every $n - k$ nonnegative elements should be greater than the absolute sum of the rest.

3.1. Null Space of Adjacency Matrices

Now let's assume that $\mathbf{A}$ is the adjacency matrix of a bipartite graph with n nodes on the left and m nodes on the right. We also assume that the graph is left d-regular. In other words $\mathbf{A}$ is a $(m \times n)$ matrix with exactly d ones in each column. First note that the following lemma holds for any such matrix.

Lemma 3.1. *Let* $\mathbf{A}^{m\times n}$ *be the adjacency matrix of a* d*-left regular bipartite graph. For any vector* $\mathbf{w}$ *in the null space of* $\mathbf{A}$*, we let* $\mathbf{w}^+$ *be the non-negative part of* $\mathbf{w}$ *and* $\mathbf{w}^-$ *be its negative portion. Then, the following is true:* $||\mathbf{w}^+||_1 = ||\mathbf{w}^-||_1$

Proof. Let $\mathbf{1} = [1, 1, ..., 1]$ be the $m \times 1$ vector of all 1's. We have:

$$\mathbf{Aw} = 0 \Rightarrow \mathbf{1Aw} = 0 \Rightarrow d\sum_{i=1}^{n} \mathbf{w}_i = 0$$

■

We are now ready to present the characterization of measurement matrices for non-negative vectors in terms of the negative support of every vector in their null space.

Theorem 3.2. *For any matrix* $\mathbf{A}^{m\times n}$ *with exactly* d *1's in each column and other entries zero, the following two statements are equivalent:*

- *Every nonnegative vector* $\mathbf{x}_0$ *with at most* k *nonzeros is a solution to (2) with* $\mathbf{y} = \mathbf{Ax}_0$.
- *Every vector* $\mathbf{w}$ *in the null space of* $\mathbf{A}$ *has at least* k *negative elements.*

Proof. We only need to show that for any $\mathbf{w} \in \mathcal{N}(\mathbf{A})$ the second statements of Theorem 3.1 and Theorem 3.2 are equivalent. Let's assume there exists a $\mathbf{w} \in \mathcal{N}(\mathbf{A})$ with less than k negative elements. We use $S^+_{\mathbf{w}}$, $S^-_{\mathbf{w}}$ and $S^0_{\mathbf{w}}$ to denote the support of positives, negatives and zeros of $\mathbf{w}$ respectively. By Lemma 3.1, $||\mathbf{w}_{S^+_{\mathbf{w}}}||_1 = ||\mathbf{w}_{S^-_{\mathbf{w}}}||_1$. Therefore any subset $S_1 \subset S^+_{\mathbf{w}} \cup S^0_{\mathbf{w}}$ that has non-empty intersection with $S^0_{\mathbf{w}}$ satisfies $||\mathbf{w}_{S^c_1}||_1 > ||\mathbf{w}_{S_1}||$. One can choose one such S_1 of

[2] $S^c = \{1, 2, ..., n\}\backslash S$. By $\mathbf{w}_S$ we mean the sub-vector of $\mathbf{w}$ constructed by those elements indexed in S.

size $n-k$ and this means there are $n-k$ non-negatives of $\mathbf{w}$ that sum to less than the absolute sum of the rest of $\mathbf{w}$. The other direction is straightforward. If any $\mathbf{w} \in \mathcal{N}(\mathbf{A})$ has k negatives, there is only one choice for $S \subset 1, 2, ..., n$ of size k, with $\mathbf{w}_{S^c} \geq 0$, and that is $S = S^-_{\mathbf{w}}$. Lemma 3.1 then guarantees $||\mathbf{w}_S||_1 = ||\mathbf{w}_{S^c}||_1$. ■

These results show how the structure of the null space of the measurement matrix is related to the recoverability of sparse vectors. Thus to achieve our primary goal which is constructing optimal sparse measurement matrices, we need to find bipartite graphs with non-negative null space properties up to a maximal sparsity (hopefully, proportional to the dimension n).We present some theorems paraphrasing the null-space property and interpreting it in terms of other properties of matrices.

3.2. Complete Rank and Natural Expansion

Before proceeding, let us consider the two following definitions, whose relation to the main topic will be shortly made apparent.

Definition 1. *For a matrix $\mathbf{A}^{m\times n}$ we define the* Complete Rank *of $\mathbf{A}$ (denoted by $\mathcal{C}r(\mathbf{A})$) to be the maximum integer r_0 with the property that every r_0 columns of $\mathbf{A}$ are linearly independent. In other words, $\mathcal{C}r(\mathbf{A}) = \min_{\mathbf{w}\in\mathcal{N}(\mathbf{A})}(|Supp(\mathbf{w})|-1)$, where by $Supp(\mathbf{w})$ we mean the support set of $\mathbf{w}$.*

Definition 2. *A left regular bipartite graph (X,Y,d) with X and Y the set of left and right vertices($|X| = n, |Y| = m$) and d the regular left degree is called a (k,ϵ) unbalanced expander if for every $S \subset X$ with $|S| \leq k$, the following holds: $|\Gamma(S)| \geq kd(1-\epsilon)$, where $\Gamma(S)$ is the set of neighbors of S in Y.*

The following lemma is connecting these two notions:

Lemma 3.2. *Every bipartite graph with adjacency matrix $\mathbf{A}$ and left degree d is a $(\mathcal{C}r(\mathbf{A}), \frac{d-1}{d})$ expander.*

Proof. Omitted for breivity. ■

A direct corollary of this theorem is that:

$$\forall S \subseteq X, |\Gamma(S)| \geq \min(|S|, \mathcal{C}r(\mathbf{A})) \tag{3}$$

The notion of complete rank is tightly related to the expansion property. It is also related to the null space characterization we are shooting for. The following theorem sheds some light on this fact.

Theorem 3.3. *If $\mathbf{A}^{m\times n}$ is the adjacency matrix of a left d-regular bipartite graph, then for every vector w in the null space of $\mathbf{A}$ the number of negative elements of w is at least $\frac{\mathcal{C}r(\mathbf{A})}{d}$.*

Proof. Let X and Y be the sets of left and right vertices of the bipartite graph corresponding to $\mathbf{A}$ (X corresponds to columnns of $\mathbf{A}$). let $S^+_{\mathbf{w}}$ be the set of vertices in X corresponding to the positive elements of $\mathbf{w}$, and $S^-_{\mathbf{w}}$ corresponding to the negative elements.[3] Let $S_{\mathbf{w}} = S^+_{\mathbf{w}} \cup S^-_{\mathbf{w}}$. The equation $\mathbf{Ax} = \mathbf{y}$ can be manifested on the graph representation of $\mathbf{A}$ with each node of Y correspond to an equation with zero R.H.S. This entails $\Gamma(S^+_{\mathbf{w}}) = \Gamma(S^-_{\mathbf{w}}) = \Gamma(S_{\mathbf{w}})$, since otherwise there is a vertex in Y connected to exactly one of $S^+_{\mathbf{w}}$ or $S^+_{\mathbf{w}}$, and its corresponding equation will not sum up to zero. On the other hand from the definition of $\mathcal{C}r(\mathbf{A})$ we must have $|S| \geq \mathcal{C}r(\mathbf{A})$. The number of edges emanating from $S^-_{\mathbf{w}}$ is $d|S^-_{\mathbf{w}}|$, which is at least as large as the number of its neighbors $|\Gamma(S^-_{\mathbf{w}})|$. Then:

$$d|S^-| \geq |\Gamma(S^-)| = |\Gamma(S)| \geq \mathcal{C}r(\mathbf{A})$$

Where the last inequality is a consequence of (3). ■

We now turn to the task of constructing adjacency matrices with complete rank proportional to dimension. Throughout this paper, all the thresholds that we achieve are asymptotical, i.e., for the regime of very large n and m.

3.3. Perturbed Expanders

When n and $m = \beta n$ are big, we are interested in constructing 0-1 matrices $A^{m\times n}$ with d(constant) 1's in each column such that $\mathcal{C}r(\mathbf{A})$ is proportional to n. Furthermore, the maximum achievable value of $\frac{\mathcal{C}r(\mathbf{A})}{nd}$ is critical. This is a very difficult question to address. However, it turns out to be much easier if we allow for a *small* perturbation of the nonzero entries of $\mathbf{A}$, as shown next.

Lemma 3.3. *For a matrix $\mathbf{A} \in \mathbb{R}^{m\times n}$ which is the adjacency matrix of a bipartite left d-regular graph, if the submatrix formed by any r_0 columns of $\mathbf{A}$ has at least r_0 nonzero rows, then it is possible to perturb nonzero entries of $\mathbf{A}$ and obtain another nonnegative matrix $\tilde{\mathbf{A}}$ through this procedure, with $\mathcal{C}r(\tilde{\mathbf{A}}) \geq r_0$. Furthermore, perturbations can be done in a way that the sum of each column remains a constant d.*

Proof. The proof is based on showing that the set of perturbations that do not guarantee $\mathcal{C}r(\tilde{\mathbf{A}}) \geq r_0$ is measure zero. We omit the proof for brevity. ■

It is worth noticing that, after modifying $\mathbf{A}$ based on perturbations described above, Theorem 3.1, Lemma 3.1 and Theorems 3.2 and 3.3 all continue to hold for this class of matrices $\tilde{\mathbf{A}}$. Therefore $\mathcal{C}r(\tilde{\mathbf{A}}) \geq r_0$ will guarantee perfect recovery of $\frac{r_0}{d}$-sparse vectors via ℓ_1-minimization. Moreover, the fact that $\mathcal{C}r(\tilde{\mathbf{A}}) \geq r_0$ can be translated back as if $\mathbf{A}$ is a

[3] We interchangeably use S and its variations to denote a set of vertices or a support set of a vector

$(r_0, \frac{d-1}{d})$ unbalanced expander graph. Therefore what we really care about is constructing $(r_0, \frac{d-1}{d})$ expanders with $\frac{r_0}{nd}$ as large as possible. In section 4, we use a probabilistic method to show that the desired $(r_0 = \mu n, \frac{d-1}{d})$ expanders exist and give thresholds on $\frac{\mu}{d}$. Before continuing, note that we are using a $1-\epsilon = \frac{1}{d}$ expansion coefficient for perfect recovery, which is very small compared to other schemes that use expanders (see, e.g., [10, 5, 6, 8, 11, 12]) and require expansion coefficients at least larger than $1-\epsilon \geq \frac{3}{4}$. $1-\epsilon = \frac{1}{d}$ is indeed the *critical* expansion coefficient. We shortly digress in a subsection to discuss this a little further.

4. EXISTENCE OF SPARSE MATRICES WITH LINEAR COMPLETE RANK

For fixed values of $n > m > r_0$ and d we are interested in the question of existence of a $(r_0, \epsilon = \frac{d-1}{d})$ expander with constant left degree d. We use the standard first moment method argument to prove the existence of such an expander for appropriate ratios of n,m and r_0 and d. The main result is given below.

Theorem 4.1. *For sufficiently large n, with $m = \beta n$ and $r_0 = \mu n$, there exists a bipartite graph with left vertex size n and right size m which is a $(r_0, \frac{d-1}{d})$ expander, if*

$$d > \frac{H(\mu) + \beta H(\frac{\mu}{\beta})}{\mu \log(\frac{\beta}{\mu})}. \tag{4}$$

Proof. (*Sketch*) Assuming that we generate a random matrix $\mathbf{A}$ by randomly generating its columns, it suffices to show that the probability that $\mathbf{A}$ has the desired expansion property is positive. For $1 \leq i_1 < i_2 < \ldots < i_r \leq n$ We denote by $E_{i_1,i_2,\ldots,i_r}$ the event that the columns of A corresponding to the numbers i_1,i_2,...,i_r have at least $n-r-1$ entire 0 rows and we have:

$$\mathbb{P}[\mathbf{A} \text{ is a } (r_0, \tfrac{d-1}{d})\text{-Exp.}] = 1 - \mathbb{P}[\mathbf{A} \text{ not a } (r_0, \tfrac{d-1}{d}\text{-Exp.})]$$

$$= 1 - \mathbb{P}[\bigcup_{d \leq r \leq r_0, 1 \leq i_1 < i_2 < \ldots < i_r} E_{i_1,i_2,\ldots,i_r}]$$

$$= 1 - \sum_{r=d}^{r_0} \binom{n}{r} \mathbb{P}[E_{1,2,\ldots,r}]$$

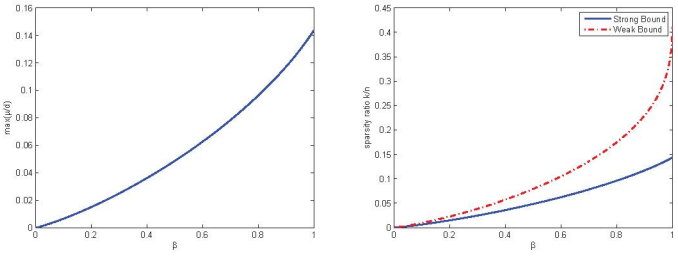
A combinatorial analysis yields the following:

$$\mathbb{P}[E_{1,2,\ldots,r}] \leq \frac{\binom{m}{r}\binom{r}{d}^r}{\binom{m}{d}^r}$$

Hence

$$\mathbb{P}[A \text{ is a } (r_0, \frac{d-1}{d})\text{-Exp.}] \geq 1 - \sum_{r=d}^{r_0} \binom{n}{r} \frac{\binom{m}{r}\binom{r}{d}^r}{\binom{m}{d}^r} \tag{5}$$

The objective is to show that this probability is positive. Equivalently, we show that for certain regimes of β, μ and d, the summation on R.H.S of (5) vanishes. To this end, we split the sum into a sub-linear summation and a summation. for $d > 2$, the sub-linear part will decays polynomially as $n \to \infty$. By applying Stirling approximations to the terms of the linear part, we find out that it decayed exponentially in n, provided (4) holds. This completes the proof. ■



(a) Strong bound. (b) Weak vs strong bound.

Fig. 1: Comparison of weak achievable bound of Section 4.1 and with the strong achievable threshold of (4) for $\frac{\mu}{d}$.

More important is the question of how big the ratio $\frac{\mu}{d}$ can be, since we earlier proved that we can recover up to $\frac{r_0}{d} = \frac{\mu}{d}n$ sparse vectors using this method. Figure 1a illustrates the maximum achievable ratio for different values of β derived from (4).

4.1. Weak bound

We are now interested in deriving conditions for recovering a specific support set S of size $k = \alpha n$, rather than obtain a worst case bound for matrices that work for all support sets. Recall that $m = \beta n$, left degree is d, and define $\gamma_1 := (1 - e^{-d\frac{\alpha}{\beta}})\beta$.

Theorem 4.2. *Define the function*

$$F(\rho_1, \rho_2) := \alpha H(\frac{\rho_1}{\alpha}) + (1-\alpha)H(\frac{\rho_2}{1-\alpha})$$
$$+\beta H(\frac{\rho_1+\rho_2}{\beta}) + d(\rho_1+\rho_2)\log(\frac{\rho_1+\rho_2}{\beta}). \tag{6}$$

For every α such that $F(\rho_1, \rho_2) < 0$ for every ρ_1, ρ_2 that satisfies $\rho_1 < \alpha, \rho_2 < 1-\alpha, \rho_1 + \rho_2 < \gamma_1$, a randomly selected subset of size $k = \alpha n$ is recoverable from a random perturbed matrix $\tilde{\mathbf{A}}$ with probability $1 - o(1)$.

The bound that results from Theorem 4.2 is plotted in Figure 1b and has been compared to the strong threshold previously achieved.

Proof. (*Sketch*) A modification of Theorem 3.1 is:

Lemma 4.1. *If $\mathbf{A}$ is the adjacency matrix of a bipartite graph with left constant degree, and if $\mathbf{x}_0$ is a fixed nonnegative vector and $\mathbf{y} = \mathbf{A}\mathbf{x}_0$, then the solution $\mathbf{x}$ of (2) will be identical*

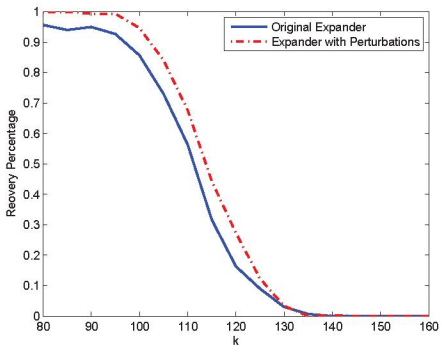


Fig. 2: Recovery percentage of ℓ_1-minimization on expanders and perturbed expanders.

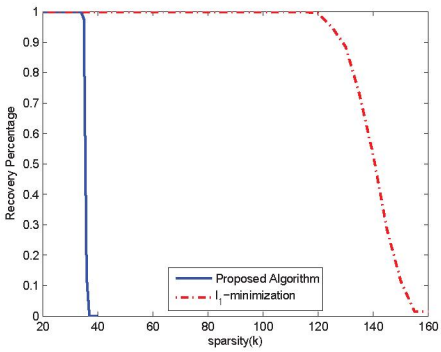


Fig. 3: Recovery percentage of Algorithm 1 Vs. ℓ_1-minimization.

to $\mathbf{x}_0$ *if and only if there exists no* $\mathbf{w}$ *in the null space of* $\mathbf{A}$ *so that* $\mathbf{w}_S^c$ *is a nonnegative vector, where* S *is the support set of* $\mathbf{x}_0$. *In other words* $\mathbf{x}_0$ *will be recoverable via L.P from* $\mathbf{A}\mathbf{x}_0$ *provided the support of* $\mathbf{x}_0$ *does not include the index set of all negative elements of a vector in the null-space of* $\mathbf{A}$.

Proof. Similar to the proof of Theorem 3.1 with S as the support of $\mathbf{x}_0$ ■

This last statement allows us to derive a combinatorial matching condition for the recovery of a vector supported on a specific subset S. We repeat the statement of the lemma:

Lemma 4.2. *Given a set* S *consider* $\Gamma(S)$ *and denote* $S_2 = \Gamma(\Gamma(S)) \setminus S$. *Let the bipartite two-hop graph of* S *be denoted by* $B_S = (S \cup S_2, \Gamma(S \cup S_2))$. *Any non-negative vector* $\mathbf{x}_0$ *supported on* S *can be recovered from* $\mathbf{y} = \tilde{\mathbf{A}}\mathbf{x}_0$ *if every subset* $S_i \subset S \cup S_2$ *of size* $|\Gamma(S) + 1|$ *has a perfect matching in* B_S.

Observe that the expectation is (asymptotically) $\mathbb{E}\Gamma(S) = (1 - e^{-d\frac{|S|}{m}})\beta n =: \gamma_1 n$ Using a standard Chernoff bound [20] it is easy to show that $\Gamma(S)$ is concentrated around its expectation:

$$\mathbb{P}[\Gamma(S) <= \mathbb{E}\Gamma(S) + \epsilon_1] > 1 - \frac{1}{n},$$

Now using the probabilistic method we can show that by randomly generating A (d-left regular) the probability that the above combinatorial matching condition holds for all subsets S_i is $1 - o(1)$, provided S is of appropriate size αn that satisfies the condition of Theorem 4.2.

■

5. FAST ALGORITHM

We now describe a fast algorithm for the recovery of sparse non-negative vectors from noiseless measurements. This algorithm relies on the minimal expansion we described in section **??**. We employ a $(kd + 1, 1 - \frac{1}{d})$ expander and perturb it as Lemma (3.3) to obtain a sparse nonnegative matrix $\tilde{\mathbf{A}}$ with $Cr(\tilde{\mathbf{A}}) \geq kd + 1$.

Algorithm 1. *Reverse Expansion Recovery*

1. *Find zeros of* $\mathbf{y}$ *and arbitrarily choose* $m - rd$ *of them and denote them by* $\mathbf{y}_1$. *Also denote by* T_1 *the index set of elements of* $\mathbf{y}_1$ *in* $\mathbf{y}$, *and by* T_2 *its complement. Wlog assume that* $\mathbf{y} = \begin{bmatrix} \mathbf{y}_1 \\ \mathbf{y}_2 \end{bmatrix}$.
2. *Locate in* X *the neighbors of the set of nodes in* Y *corresponding to* T_1, *name the set* S_1 *and name the set of their complement nodes in* X *by* S_2.
3. *Identify the sub-matrix of* $\tilde{\mathbf{A}}$ *that represents the edges emanating from* S_2 *to* T_2. *Call this sub-matrix* $\tilde{\mathbf{A}}_2$. *Columns of* $\tilde{\mathbf{A}}_2$ *correspond to nodes in* S_2, *and its rows correspond to the nodes in* T_2.
4. *Set* $\hat{\mathbf{x}}_{S_1} = 0$ *and compute* $\hat{\mathbf{x}}_{S_2} = \tilde{\mathbf{A}}_2^{\dagger}\mathbf{y}_2$, *where* $\mathbf{A}^{\dagger}$ *is the pseudo-inverse of* $\mathbf{A}$ *defined by* $\mathbf{A}^{\dagger} = (\mathbf{A}^t\mathbf{A})^{-1}\mathbf{A}^t$. *Declare* $\hat{\mathbf{x}}$ *as the output.*

The algorithm begins with identifying a big zero portion of the output and locating their corresponding nodes in Y. In the next step, neighbors of these nodes are found in X and these two giant sets of nodes are eliminated from X and Y. Having done this, we are left with a much smaller system of linear equations, which turns out to be *over-determined*, and therefore our problem reduces to solving linear equations. Following theorem provides theoretical guarantees for the validity of this algorithm.

Theorem 5.1. *Validity of Algorithm*

If $\mathbf{x}$ *is a* k*-sparse non-negative vector and* $\tilde{\mathbf{A}}$ *is a perturbed* $(kd + 1, 1 - \frac{1}{d})$ *expander with* $Cr(\mathbf{A}) \geq kd + 1$ *then:*

1. $\mathbf{y}$ *is* kd*-sparse and therefore has at least* $m - kd$ *zeros.*
2. $|S_2| \leq |T_2|$ *and* $\tilde{\mathbf{A}}_2$ *is tall (or square) and therefore full rank.*
3. $\hat{\mathbf{x}} = \mathbf{x}$

Proof.

1. Trivial.

2. Suppose $|S_2| > |T_1| = rd$. select an arbitrary subset of ${S_2}' \subseteq S_2$ of size $kd+1$. Because of the expansion property: $|\Gamma({S_2}')| \geq kd+1 > |T_1|$. But $\Gamma({S_2}')$ is in T_2 and this is a contradiction.
3. If any entry in $\mathbf{x}_{S_1}$ is positive, then the equations corresponding to its neighbors in T_1 are not zero . This is in contradiction with the choice of T_1. So $\mathbf{x}_{S_1} = 0 = \hat{\mathbf{x}}_{S_1}$. Also since $\tilde{\mathbf{A}}_2 \mathbf{x}_{S_2} = \mathbf{y}_2$, $\tilde{\mathbf{A}}_2 \hat{\mathbf{x}}_{S_2} = \mathbf{y}_2$ and $\tilde{\mathbf{A}}_2$ is full rank we conclude that $\mathbf{x}_{S_2} = \hat{\mathbf{x}}_{S_2}$.

■

6. EXPERIMENTAL EVALUATION

We generated a random $m \times n$ matrix $\mathbf{A}$ with $n = 2m = 500$, and $d = 3$ 1's in each column. We then multiplied random sparse vectors with different sparsity levels by $\mathbf{A}$, and tried recovering them via the linear programming (2). Next, we added the perturbations described in section 3 to $\mathbf{A}$ and applied the same sparse vectors to compare the recovery percentages in the two cases. This process was repeated for a few generations of $\mathbf{A}$ and the best of the improvements we obtained is illustrated in Figure 2.

7. REFERENCES

[1] David Donoho,"High-dimensional centrally symmetric polytopes with neighborliness proportional to dimension," *Discrete and Computational Geometry*, 102(27),pp. 617-652 2006, Springer.

[2] David Donoho and Jared Tanner "Neighborlyness of randomly-projected simplices in high dimensions," *Proc. National Academy of Sciences*, 102(27), pp. 9452-9457, 2005.

[3] E.Candes, I.Romberg and T.Tao "Robust Uncertainty Principles: Exact Signal Reconstruction from Highly Incomplete Frequency Information," *IEEE Trans. on Info. Theory*, 52(2) pp. 489-509, Feb. 2006.

[4] David L.Donoho and Jared Tanner "Sparse Nonnegative Solution of Underdetermined Linear Equations by Linear Programming", *Proceedings of the National Academy of Sciences*, Vol. 102(27) (2005) 9446-9451.

[5] A. Gilbert, P. Indyk, H. Karloff, and M. Strauss "Combining geometry and combinatorics: a unified approach to sparse signal recovery". Manuscript, 2007.

[6] P. Indyk and M. Ruzic "Near-Optimal Sparse Recovery in the L_1 norm", 49th Symposium on Foundations of Computer Science (FOCS), 2008.

[7] J. Tropp and A. Gilbert, "Signal recovery from random measurements via orthogonal matching pursuit". IEEE Trans. on Information Theory, 53(12) pp. 4655-4666, December 2007.

[8] R. Berinde and P. Indyk "Sparse recovery using sparse matrices", *Computer Science and Artificial Intelligence Laboratory Technical Report*, MIT-CSAIL-TR-2008-001,January 10, 2008.

[9] Alfred M. Bruckstein, Michael Elad, and Michael Zibulevsky, "A non-negative and sparse enough solution of an underdetermined linear system of equations is unique", Preprint, 2007.

[10] Weiyu Xu and Babak Hassibi "Efficient compressive sensing with determinstic guarantees using expander graphs", *IEEE Information Theory Workshop*, Lake Tahoe, September 2007.

[11] Weiyu Xu and Babak Hassibi "Further Results on Performance Analysis for Compressive Sensing Using Expander Graphs", Signals, Systems and Computers, 2007. ACSSC 2007. Conference Record of the Forty-First Asilomar Conference on 4-7 Nov. 2007 Page(s):621 - 625.

[12] Sina Jafarpour, Weiyu Xu, Babak Hassibi, and Robert Calderbank "Efficient compressed sensing using high-quality expander graphs". Preprint, 2008).

[13] Weiyu Xu and Babak Hassibi "Compressed sensing over the Grassmann manifold: A unified analytical framework". Preprint, 2008.

[14] P. Indyk Explicit constructions for compressed sensing of sparse signals, SODA, 2008.

[15] Farzad Parvaresh, Haris Vikalo, Sidhant Misra, and Babak Hassibi "Recovering sparse signals using sparse measurement matrices in compressed DNA microarrays"*IEEE Journal of Selected Topics In Signal Proc.*, Vol.2, Issue 3, pp. 275-285 06-2008.

[16] Mihailo Stojnic, Weiyu Xu, and Babak Hassibi "Compressed sensing - probabilistic analysis of a null-space characterization"*IEEE International Conference on Acoustic, Speech and Signal Processing, ICASSP 2008*

[17] V. Guruswami, J.R. Lee, A. Wigderson "Euclidean sections of L_1^n with sublinear randomness and error-correction over the reals", RANDOM 2008.

[18] V. Guruswami, J.R. Lee, A. Razborov "Almost Euclidean subspaces of L_1^n via expander codes", SODA 2008.

[19] A. Feuer and A. Nemirovski "On sparse representation in pairs of bases" *IEEE Transactions on Information Theory* 49(6): 1579-1581 (2003).

[20] Michael Mitzenmacher, Eli Upfal "Probability and Computing: Randomized Algorithms and Probabilistic Analysis". Cambridge University Press, 2005.

[21] N. Linial and I. Novik "How neighborly can a centrally symmetric polytope be?" *Discrete and Computational Geometry,* 36(2006) 273-281.

[22] M. Sipser and D.A. Spielman, "Expander Codes", IEEE Transactions on Information Theory, 1996.

[23] Y. Zhang "When is missing data recoverabl". Rice CAAM Department Technical Report TR05-15, 2005, available online at http://www.dsp.ece.rice.edu/cs/.